Biometrika (2025), **112**, 4, asaf072
https://doi.org/10.1093/biomet/asaf072
Advance Access publication 16 October 2025

On testing Kronecker product structure in tensor factor models

BY ZETAI CEN

School of Mathematics, University of Bristol, Woodland Road, Bristol BS8 1UG, U.K.
zetai.cen@bristol.ac.uk

AND CLIFFORD LAM

Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, U.K.
c.lam2@lse.ac.uk

SUMMARY

We propose a test for the Kronecker product structure of a factor loading matrix implied by a tensor factor model with Tucker decomposition in the common component. By defining a Kronecker product structure set, we determine whether a tensor time series has a Kronecker product structure, equivalent to its ability to decompose the series according to a tensor factor model. Our test is built on analysing and comparing the residuals from fitting a full tensor factor model, and the residuals from fitting a factor model on a reshaped version of the data. In the most extreme case, the reshaping is the vectorization of the tensor data, and the factor loading matrix in such a case can be general if there is no Kronecker product structure present. Our test is also generalized to the Khatri–Rao product structure in a tensor factor model with canonical polyadic decomposition. Theoretical results are developed through asymptotic normality results on estimated residuals. Numerical experiments suggest that the size of the tests approaches the pre-set nominal value as the sample size or the order of the tensor increases, while the power increases with mode dimensions and the number of combined modes. We demonstrate our tests through extensive real data examples.

Some key words: Factor-structured idiosyncratic error; Tensor refold; Tensor reshape; Weak factor.

1. INTRODUCTION

With rapid advances in information technology, high-dimensional time series data observed in tensor form are becoming increasingly available for analysis in fields such as finance, economics, bioinformatics and computer science, to name but a few areas. In many cases, low-rank structures observed in the tensor time series can be exploited, facilitating analysis and interpretation. The most commonly used devices are the canonical polyadic (CP) decomposition and the multilinear/Tucker decomposition of a tensor, leading to CP tensor factor models (Chang et al., 2023; Han et al., 2024b) and Tucker tensor factor models (Chen et al., 2022, 2024; Chen & Lam, 2024; Han et al., 2024a) for tensor time series factor models, respectively. While tensor time series can be transformed back to vector time series through vectorization and be analysed using traditional factor models for vector time series, the tensor structure of the data is lost and hence any corresponding interpretations from it. Moreover, vectorization significantly increases the dimension of the factor loading matrix relative to the sample size, potentially leading to less accurate estimation and inferences (Chen & Lam, 2024).

However, a tensor factor model comes with its assumptions. In using the Tucker decomposition in particular, a tensor factor model assumes that the factor loading matrix for the vectorized data is the Kronecker product of lower-dimensional factor loading matrices. For instance, suppose that a mean-zero matrix $Y_t \in \mathbb{R}^{d_1 \times d_2}$ is observed at each $t = 1, \ldots, T$. Consider a matrix factor model (first studied by Wang et al., 2019, and further extended/analysed by, e.g., Yu et al., 2022a and Chen & Fan, 2023) of the form

$$Y_t = A_1 F_t A_2^{\mathrm{T}} + E_t, \tag{1}$$

where $F_t \in \mathbb{R}^{r_1 \times r_2}$ is the core factor, $A_k \in \mathbb{R}^{d_k \times r_k}$ is the mode-k factor loading matrix, i.e., A_1 and A_2 are respectively the row and column loading matrices, and E_t is the noise. The vectorization of (1) is

$$\mathrm{vec}(Y_t) = (A_2 \otimes A_1)\,\mathrm{vec}(F_t) + \mathrm{vec}(E_t) \equiv A_V\,\mathrm{vec}(F_t) + \mathrm{vec}(E_t), \tag{2}$$

where $A_V = A_2 \otimes A_1$, which is a vector factor model for the time series data $\{\mathrm{vec}(Y_t)\}$ with factor loading matrix A_V. Clearly, the implicit assumption of a Kronecker product structure for A_V when using a matrix factor model for matrix-valued time series data should be the first aspect to check before applying such a factor model.

Motivated by this simple example, we propose a test in this paper to primarily assess the Kronecker product structure of the factor loading matrix implied in the vectorized data when using a Tucker tensor factor model (TFM), and extend it to higher-order tensors. He et al. (2023) also noted this implicit assumption in a Tucker matrix factor model and proposed testing the boundary cases where each column (or row, respectively) of the data follows a factor model with a common factor loading matrix, but possibly distinct factors, or where the entire matrix is simply pure noise. Model (2) with a general A_V also implies a vector factor model with potentially different factor loading matrices for each column (or row, respectively) of the data, but with shared factors. To explore the data as a matrix, connectedness through having a set of shared common factors rather than having the same factor loading matrix with all distinct factors is more meaningful. Practically, (2) is an alternative model that fits the data more easily than the boundary cases of He et al. (2023), since the data still follow a more general factor model, though the implied Kronecker product structure in the factor loading matrix A_V is lost. This comes as no surprise, then, as in all of the tests in He et al. (2023) for their real data analyses, they cannot reject the null hypothesis of a matrix factor model. An easier alternative, such as (2) with just a general A_V, can provide a more critical test for the null hypothesis of a matrix factor model. See our portfolio return and macroeconomic indices examples in § 5.2 for cases where our test can reject the null hypothesis of a matrix factor model, while He et al. (2023) cannot.

We also stress that our model is fundamentally different from those used to test for a Kronecker product structure in the covariance matrix of the data. For example, Yu et al. (2022b) and Guggenberger et al. (2023) both proposed tests for the Kronecker product structure of the covariance matrix of vectorized matrix data. For model (1), even in the simplest hypothetical case where E_t and F_t are independent and F_t contains independent standard normal random variables, we have

$$\text{cov}\{\text{vec}(Y_t)\} = A_2A_2^{\mathrm{T}} \otimes A_1A_1^{\mathrm{T}} + \text{cov}\{\text{vec}(E_t)\},$$

so the covariance matrix is never exactly of Kronecker product structure because of E_t. Moreover, even with $E_t = 0$, both $A_1A_1^{\mathrm{T}}$ and $A_2A_2^{\mathrm{T}}$ are of low rank, which is different from the full-rank component matrices in the two papers mentioned above.

Our contributions in this paper are three-fold. First, as a first in the literature, we propose a test to assess a direct Tucker TFM against the alternative of a TFM in which the Kronecker product structure lost in some of its factor loading matrices. As shown in § 3, for higher-order tensors, testing against a TFM of a time series can be conducted on a TFM for the reshaped data, but not necessarily on the vectorized data. Moreover, our test can be easily generalized to test the Khatri–Rao product structure in CP TFM; see Appendix B within the Supplementary Material and the real data analysis in § 5.2. All these features give rise to flexibility and, in fact, statistical power in practical situations. Second, our analysis allows for weak factors, and our theoretical results explicitly characterize the rates of convergence. Last but not least, as a useful by-product, we developed tensor reshaping theorems that can be insightful and useful in their own right.

2. Notation and tensor reshaping

2.1. *Notation*

Throughout this paper, we use a lowercase letter, capital letter and calligraphic letter, i.e., $x, X, \mathcal{X}$, to denote a scalar or a vector, a scalar or a matrix and a tensor, respectively. We also use $x_i, X_{ij}, X_{i\cdot}, X_{\cdot i}$ to respectively denote the ith element of a vector x, the (i,j)th element of X, the ith row vector (as a column vector) of X and the ith column vector of X. We use $\otimes$ to represent the Kronecker product, $\odot$ the Khatri–Rao product, $*$ the Hadamard product and $\circ$ the tensor outer product; see the Supplementary Material for more details. By convention, the total Kronecker or Khatri–Rao product over an index set is computed in descending index order. We use $a \asymp b$ to denote $a = O(b)$ and $b = O(a)$. Hereafter, given a positive integer m, define $[m] = \{1, 2, \ldots, m\}$. The ith largest eigenvalue of a matrix X is denoted by $\lambda_i(X)$. We denote by $\text{tr}(X)$ the trace of X, and by X^{T} the transpose of X. Define $d = \prod_{k=1}^K d_k$, $d_{\text{-}k} = d/d_k$, $r = \prod_{k=1}^K r_k$ and $r_{\text{-}k} = r/r_k$.

Sets are also denoted by calligraphic letters. For a given set $\mathcal{A}$, we denote by $|\mathcal{A}|$ and $\mathcal{A}_i$ its cardinality and the ith element, respectively. We use $\|\cdot\|$ to denote the spectral norm of a matrix or the L_2 norm of a vector, and $\|\cdot\|_F$ to denote the Frobenius norm of a matrix. We use $\|\cdot\|_{\max}$ to denote the maximum absolute value of the elements in a vector, a matrix or a tensor. The notation $\|\cdot\|_1$ and $\|\cdot\|_\infty$ denote the L_1 and L_∞ norms of a matrix, respectively, defined by $\|X\|_1 = \max_j \sum_i |X_{ij}|$ and $\|X\|_\infty = \max_i \sum_j |X_{ij}|$. Without loss of generality, we always assume that the eigenvalues of a matrix are arranged in descending order, as are their corresponding eigenvectors.

For the rest of this section, we briefly introduce the notation and operations for tensor data. For more details on tensor manipulations and decompositions, we refer the reader to the Supplementary Material. A multi-dimensional array with K dimensions is an *order-K* tensor, with its kth dimension termed *mode k*. For an order-K tensor $\mathcal{X} = (X_{i_1,\ldots,i_K}) \in \mathbb{R}^{d_1\times\cdots\times d_K}$, a column vector $(X_{i_1,\ldots,i_{k-1},i,i_{k+1},\ldots,i_K})_{i\in[d_k]}$ represents a *mode-k fibre* for tensor $\mathcal{X}$. We denote by $\text{mat}_k(\mathcal{X}) \in \mathbb{R}^{d_k\times d_{\text{-}k}}$ (or sometimes $X_{(k)}$ for convenience) the *mode-k unfolding/matricization* of a tensor, defined by placing all mode-k fibres into a matrix. We denote by $\mathcal{X}\times_k A$ the *mode-k product* of a tensor $\mathcal{X}$ with a matrix A, defined by

$$\text{mat}_k(\mathcal{X}\times_k A) = A\,\text{mat}_k(\mathcal{X}).$$

We use the notation $\text{vec}(\cdot)$ to denote the vectorization of a matrix or the vectorization of the mode-1 unfolding of a tensor. The *refolding/tensorization* of a vector $x \in \mathbb{R}^{d_1,\ldots,d_K}$ on $\{d_1,\ldots,d_K\}$ is defined to be an order-K tensor $\text{fold}(x,\{d_1,\ldots,d_K\}) \in \mathbb{R}^{d_1\times\cdots\times d_K}$ such that $x = \text{vec}[\text{fold}(x,\{d_1,\ldots,d_K\})]$. The *refolding/tensorization* of a matrix $X \in \mathbb{R}^{d_k\times d_{\text{-}k}}$ on $\{d_1,\ldots,d_K\}$ along mode k is defined to be $\text{fold}_k(X,\{d_1,\ldots,d_K\}) \in \mathbb{R}^{d_1\times\cdots\times d_K}$ such that $X = \text{mat}_k[\text{fold}_k(X,\{d_1,\ldots,d_K\})]$.

2.2. *Introduction to tensor reshaping*

In this subsection, we introduce tensor reshaping. Given an order-K tensor $\mathcal{X} \in \mathbb{R}^{d_1\times\cdots\times d_K}$ and a set with ordered, strictly ascending elements $\{a_1,\ldots,a_\ell\} \subseteq [K]$, the $\text{reshape}(\cdot,\cdot)$ operator is defined as follows.

If $\ell = 1$,

$$\text{reshape}(\mathcal{X},\{a_1\}) = \text{fold}_K(\text{mat}_{a_1}(\mathcal{X}),\{d_1,\ldots,d_{a_1-1},d_{a_1+1},\ldots,d_K,d_{a_1}\});$$

if $\ell = 2$,

$$\begin{aligned}&\text{reshape}(\mathcal{X},\{a_1,a_2\})\\&\quad= \text{fold}_{K-1}(\mathcal{X}_{a_1\sim a_2},\{d_1,\ldots,d_{a_1-1},d_{a_1+1},\ldots,d_{a_2-1},d_{a_2+1},\ldots,d_K,d_{a_1}d_{a_2}\}),\end{aligned}$$

where

$$\mathcal{X}_{a_1\sim a_2} = \begin{pmatrix}\text{mat}_{a_1}[\text{fold}(\text{mat}_{a_2}(\mathcal{X})_{1\cdot},\{d_1,\ldots,d_{a_2-1},d_{a_2+1},\ldots,d_K\})]\\ \cdots\\ \text{mat}_{a_1}[\text{fold}(\text{mat}_{a_2}(\mathcal{X})_{d_{a_2}\cdot},\{d_1,\ldots,d_{a_2-1},d_{a_2+1},\ldots,d_K\})]\end{pmatrix};$$

if $\ell \geqslant 3$,

$$\begin{aligned}&\text{reshape}(\mathcal{X},\{a_1,\ldots,a_\ell\})\\&\quad= \text{reshape}[\text{reshape}(\mathcal{X},\{a_{\ell-1},a_\ell\}),\{a_1,\ldots,a_{\ell-2},K-1\}].\end{aligned}$$

Hence, reshaping an order-K tensor along $\{a_1,\ldots,a_\ell\}$ results in an order-$(K-\ell+1)$ tensor. A heuristic view of $\text{reshape}(\mathcal{X},\{a_1,\ldots,a_\ell\})$ is that all modes of $\mathcal{X}$ with indices $\{a_1,\ldots,a_\ell\}$ are merged into a single mode acting as the last mode as a result. One may recover $\mathcal{X}$ from $\text{reshape}(\mathcal{X},\{a_1,\ldots,a_\ell\})$, given the original dimension of $\mathcal{X}$ and $\{a_1,\ldots,a_\ell\}$. To help the reader understand the reshape operator, Fig. 1 is presented as a visualization, and we refer the reader to the Supplementary Material for more details on tensor reshaping.

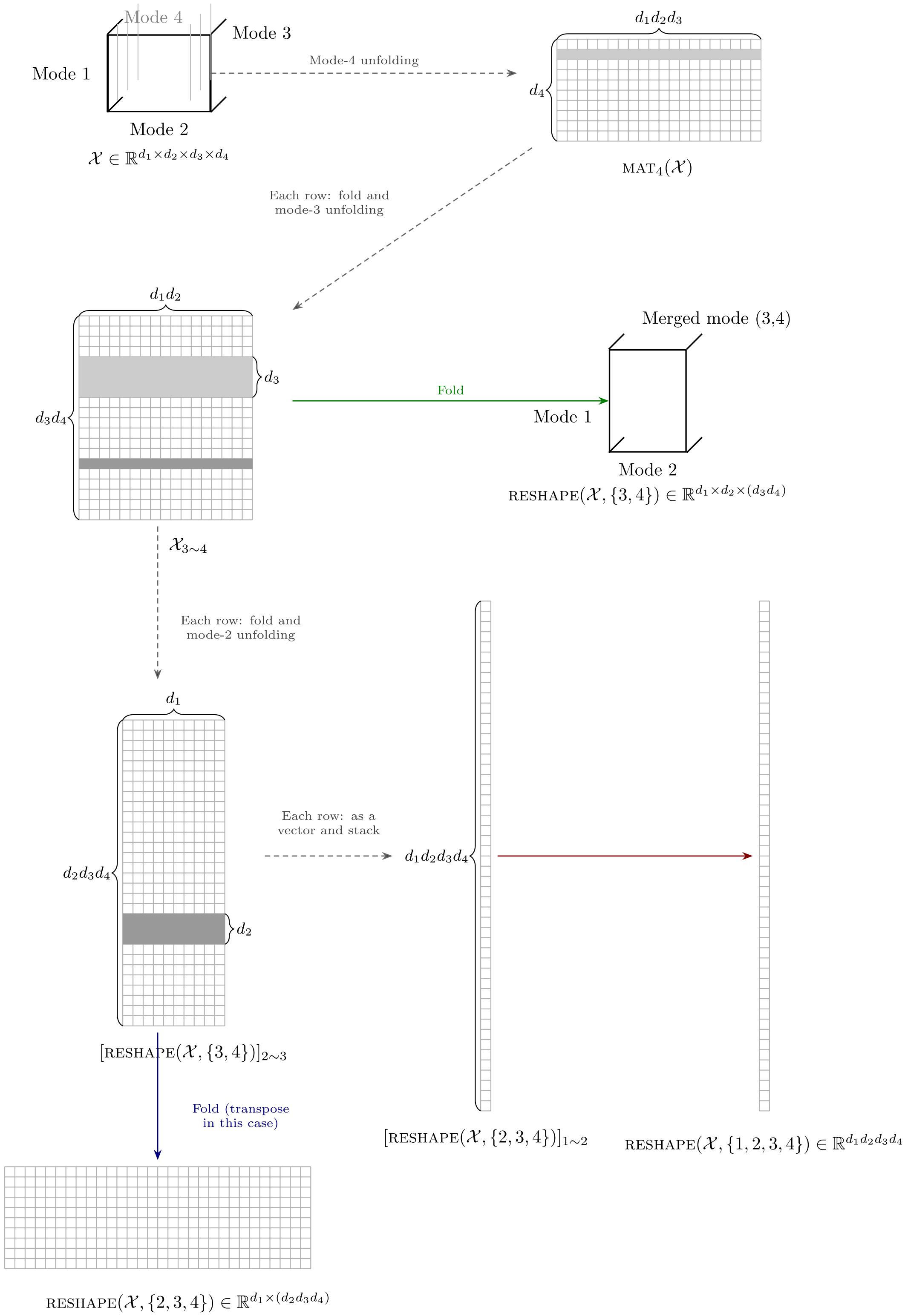

Fig. 1. Illustration of the reshape operator for an order-4 tensor $\mathcal{X}$ along $\mathcal{A}$. The last step in the reshape with $\mathcal{A} = \{3, 4\}, \{1, 2, 3, 4\}$ and $\{2, 3, 4\}$ is respectively denoted by the first horizontal solid arrow, the second horizontal solid arrow, and the vertical solid arrow.

3. FACTOR MODEL AND TESTING ITS KRONECKER PRODUCT STRUCTURE

3.1. *Factor model, Kronecker and Khatri–Rao product structures*

This subsection introduces the concept of Tucker TFM and CP TFM with Kronecker and Khatri–Rao product structures, respectively, and establishes the technical details for the testing problem. For a more integrated reading experience, the reader can go straight to § 3.2, where equations and terms are referred back to § 3.1 whenever necessary. For the details on testing the Khatri–Rao product structure for CP TFM, we refer the reader to Appendix B in the Supplementary Material. We begin by introducing sets that facilitate describing our models.

DEFINITION 1 (KRONECKER AND KHATRI–RAO PRODUCT STRUCTURE SETS). *Given an ordered set of positive integers* $\{b_1, \ldots, b_\kappa\}$*, the Kronecker product structure set is defined as*

$$\mathcal{K}_{b_1\times\cdots\times b_\kappa} = \{A \mid A = A_\kappa \otimes \cdots \otimes A_1 \text{ such that, for each } j \in [\kappa], \\ A_j \in \mathbb{R}^{b_j \times u_j} \text{ has rank } u_j \ll b_j,\ \|A_{j,\cdot i}\|^2 \asymp b_j^{\delta_{j,i}},\ \delta_{j,i} \in (0, 1]\}.$$

The Khatri–Rao product structure set, denoted $\mathcal{KR}_{b_1\times\cdots\times b_\kappa}$*, is defined similarly except* $\otimes$ *is replaced by* $\odot$*, and* $u_1 = \cdots = u_\kappa$*.*

The Kronecker product structure set defined in Definition 1 characterizes the factor loading matrix in the Tucker TFM, while the Khatri–Rao product structure set characterizes that of the CP TFM. Requiring $\delta_{j,i} > 0$ is to ensure a certain factor strength in each loading matrix. See Assumptions 1 and 2 in § 4.1 below for the technical details. The form of factor models is depicted below, featuring either a Kronecker or Khatri–Rao product structure.

DEFINITION 2 (FACTOR MODELS AND PRODUCT STRUCTURES). *Given a time series of mean-zero order-K tensors* $\mathcal{Y}_t \in \mathbb{R}^{d_1\times\cdots\times d_K}$ *for* $t \in [T]$ *and a set with ordered, ascending elements* $\mathcal{A} = \{a_1, \ldots, a_\ell\} \subseteq [K]$*, we say that* $\{\mathcal{Y}_t\}$ *follows a Tucker TFM along* $\mathcal{A}$ *if, for* $t \in [T]$*,*

$$\text{reshape}(\mathcal{Y}_t, \mathcal{A}) = \mathcal{C}_{\text{reshape},t} + \mathcal{E}_{\text{reshape},t} = \mathcal{F}_{\text{reshape},t} \times_{j=1}^{K-\ell+1} A_{\text{reshape},j} + \mathcal{E}_{\text{reshape},t}, \tag{3}$$

where $\text{reshape}(\mathcal{Y}_t, \mathcal{A}) \in \mathbb{R}^{p_1\times\cdots\times p_{K-\ell+1}}$ *(for some* $p_1, \ldots, p_{K-\ell+1}$*) is the order-*$(K-\ell+1)$ *tensor obtained by reshaping* $\mathcal{Y}_t$ *along* $\mathcal{A}$*; the common component* $\mathcal{C}_{\text{reshape},t}$ *consists of the core factor* $\mathcal{F}_{\text{reshape},t} \in \mathbb{R}^{\pi_1\times\cdots\times\pi_{K-\ell+1}}$ *and loading matrices* $A_{\text{reshape},j} \in \mathbb{R}^{p_j\times\pi_j}$*, with rank* $\pi_j \ll p_j$ *for* $j \in [K-\ell+1]$*; and* $\mathcal{E}_{\text{reshape},t}$ *is the noise. We also make the following classifications.*

(i) *The series* $\{\mathcal{Y}_t\}$ *has a Kronecker product structure if* $A_{\text{reshape},K-\ell+1} \in \mathcal{K}_{d_{a_1}\times\cdots\times d_{a_\ell}}$.
(ii) *The series* $\{\mathcal{Y}_t\}$ *has no Kronecker product structure along* $\mathcal{A}$ *if* $A_{\text{reshape},K-\ell+1} \notin \mathcal{K}_{d_{a_1}\times\cdots\times d_{a_\ell}}$.

Furthermore, we say that $\{\mathcal{Y}_t\}$ *follows a CP TFM along* $\mathcal{A}$ *if the core factor* $\mathcal{F}_{\text{reshape},t}$ *is diagonal such that* $\pi_1 = \cdots = \pi_{K-\ell+1}$*, and* $(\mathcal{F}_{\text{reshape},t})_{i_1,\ldots,i_{K-\ell+1}} \neq 0$ *only if* $i_1 = \cdots = i_{K-\ell+1}$*. The classifications are similarly defined for a CP TFM, except the Kronecker product structure is replaced by the Khatri–Rao product structure and* $\mathcal{K}_{d_{a_1}\times\cdots\times d_{a_\ell}}$ *by* $\mathcal{KR}_{d_{a_1}\times\cdots\times d_{a_\ell}}$.

Definition 2 formally defines two forms of factor models, where testing the Kronecker product structure for a Tucker TFM is the main focus of this paper. Our test can also be directly extended to testing the Khatri–Rao product structure for a CP TFM as a by-product; see the discussion in § 6 below and Appendix B in the Supplementary Material.

A key piece of information in Definition 2(i) is that if the Kronecker product structure holds along some $\mathcal{A}$ then it holds along any $\mathcal{A}$; see the discussion below Theorem 1 for details. If $\ell = 1$ in Definition 2, i.e., $\mathcal{A}$ contains only one element (representing the mode index), for each order-K tensor $\mathcal{Y}_t$, $\text{reshape}(\mathcal{Y}_t, \{a_1\})$ is the order-K tensor constructed from $\mathcal{Y}_t$ by treating mode a_1 as mode K. Hence, a Tucker TFM of $\mathcal{Y}_t$ along $\{a_1\}$ reduces to the conventional Tucker TFM (Chen et al., 2022; Barigozzi et al., 2025) of $\mathcal{Y}_t$, but with the mode indices changed. For instance, we may read (3) along $\mathcal{A} = \{K\}$ as

$$\mathcal{Y}_t = \mathcal{C}_{\text{reshape},t} + \mathcal{E}_{\text{reshape},t} = \mathcal{F}_{\text{reshape},t} \times_1 A_{\text{reshape},1} \times_2 \cdots \times_K A_{\text{reshape},K} + \mathcal{E}_{\text{reshape},t}.$$

From the above, Definition 2(i) automatically describes $\{\mathcal{Y}_t\}$ if $\ell = 1$, and the Kronecker product structure is only nontrivial for $\ell \geqslant 2$ (hence, $K \geqslant 2$). To demystify Definition 2.1, we next present Theorem 1, which, as a first in the literature, explicitly states the equivalence of the Tucker TFM under tensor reshaping.

Theorem 1 (Tensor reshaping I). *Using the notation in Definition 2, $\{\mathcal{Y}_t\}$ following a Tucker TFM along any given $\mathcal{A} = \{a_1, \ldots, a_\ell\} \subseteq [K]$, as in (3), with a Kronecker product structure, is equivalent to $\{\mathcal{Y}_t\}$ following a Tucker TFM such that*

$$\mathcal{Y}_t = \mathcal{C}_t + \mathcal{E}_t = \mathcal{F}_t \times_1 A_1 \times_2 \cdots \times_K A_K + \mathcal{E}_t, \tag{4}$$

where $\mathcal{C}_t$ is the common component, $\mathcal{F}_t \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ is the core factor, each $A_k \in \mathbb{R}^{d_k \times r_k}$ with $r_k \ll d_k$ is the mode-k loading matrix and $\mathcal{E}_t$ is the noise. More importantly, with the definition $\mathcal{A}^* = [K] \setminus \mathcal{A}$, we have

$$\mathcal{F}_{\text{reshape},t} = \text{reshape}(\mathcal{F}_t, \mathcal{A}), \qquad \mathcal{E}_{\text{reshape},t} = \text{reshape}(\mathcal{E}_t, \mathcal{A}),$$
$$A_{\text{reshape},K-\ell+1} = \bigotimes_{i \in \mathcal{A}} A_i, \qquad A_{\text{reshape},j} = A_{\mathcal{A}_j^*} \quad (j \in [K-\ell]).$$

Moreover, (4) uniquely determines (3), and (3) determines (4) up to an arbitrary set $\{A_i\}_{i \in \mathcal{A}}$.

Using the same notation as above, $\{\mathcal{Y}_t\}$ following a CP TFM along $\mathcal{A}$ with a Khatri–Rao product structure is equivalent to $\{\mathcal{Y}_t\}$ following a CP TFM as in (4) with diagonal $\mathcal{F}_t$. In this case, the reshaped factor structure is the same as above except that $A_{\text{reshape},K-\ell+1} = \bigodot_{i \in \mathcal{A}} A_i$ and $\mathcal{F}_{\text{reshape},t}$ is the order-$(K-\ell+1)$ diagonal tensor with the same diagonal entries as $\mathcal{F}_t$.

Theorem 1 reveals that a Tucker TFM (CP TFM) on $\{\mathcal{Y}_t\}$ along any $\mathcal{A}$ with Kronecker product structure (Khatri–Rao product structure) in Definition 2 is preserved under the reshape operator. This forms the foundation for the hypothesis test design later. The identifications of (3) and (4) are relegated to the Supplementary Material.

3.2. *A test on the Kronecker product structure*

For testing of the Khatri–Rao product structure for the CP TFM, we refer the reader to Appendix B in the Supplementary Material. The testing problem on the Kronecker product structure is formally defined in this subsection, with an example on an order-2 tensor (i.e., a matrix) at the end. For each $t \in [T]$, we observe a mean-zero order-K tensor $\mathcal{Y}_t \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ with $K \geqslant 2$ (otherwise, the test is trivial, as explained in § 3.1). Without loss of generality, let $v < K$ be a given positive integer and define $\mathcal{A} = \{v, \ldots, K\}$, which contains the mode indices along which the Kronecker product structure might be lost;

see the alternative hypothesis H_1 below. Suppose that $\{\mathcal{Y}_t\}$ follows a Tucker TFM along $\mathcal{A}$ as in Definition 2, with the notation therein except that we now read (3) as

$$\text{reshape}(\mathcal{Y}_t, \mathcal{A}) = \mathcal{C}_{\text{reshape},t} + \mathcal{E}_{\text{reshape},t} = \mathcal{F}_{\text{reshape},t} \mathop{\times}_{k=1}^{v-1} A_k \mathop{\times}_{v} A_V + \mathcal{E}_{\text{reshape},\, t}, \tag{5}$$

where $A_j \in \mathbb{R}^{d_j \times r_j}$ for $j \in [v-1]$ (if $v > 1$) and $A_V \in \mathbb{R}^{d_V \times r_V}$ with $d_V = \prod_{i=v}^{K} d_i$. Essentially, the order-v tensor reshape$(\mathcal{Y}_t, \mathcal{A})$ follows a Tucker TFM. The set $\{r_1, \ldots, r_{v-1}, r_V\}$ is assumed known and any consistent estimators (e.g., Han et al., 2022; Chen & Lam, 2024) can be used in practice. With $\mathcal{K}_{d_v \times \cdots \times d_K}$ defined in Definition 1, we consider a hypothesis test as follows:

$$H_0: \{\mathcal{Y}_t\} \text{ has a Kronecker product structure, i.e., } A_V \in \mathcal{K}_{d_v \times \cdots \times d_K},$$
$$H_1: \{\mathcal{Y}_t\} \text{ has no Kronecker product structure along } \mathcal{A}, \text{ i.e., } A_V \notin \mathcal{K}_{d_v \times \cdots \times d_K}. \tag{6}$$

Besides the complexity of being a composite testing problem, the difficulty with hypotheses (6) is compounded by the fact that $\mathcal{Y}_t$ under the alternative has no explicit form without reshaping along $\mathcal{A}$. Fortunately, the factor structure in (5) is stable under both hypotheses. That is, the estimation of $\{\mathcal{F}_{\text{reshape},t}, A_1, \ldots, A_{v-1}, A_V\}$ is always feasible. In particular, thanks to Theorem 1, under H_0, we have

$$\mathcal{Y}_t = \mathcal{C}_t + \mathcal{E}_t = \mathcal{F}_t \mathop{\times}_{1} A_1 \mathop{\times}_{2} \cdots \mathop{\times}_{v-1} A_{v-1} \mathop{\times}_{v} A_v \mathop{\times}_{v+1} \cdots \mathop{\times}_{K} A_K + \mathcal{E}_t, \tag{7}$$

where $A_k \in \mathbb{R}^{d_k \times r_k}$ for $k \in [K]$ (hence, the first $v-1$ loading matrices are exactly those in (5)),

$$\text{reshape}(\mathcal{F}_t, \mathcal{A}) = \mathcal{F}_{\text{reshape},\, t}, \qquad \text{reshape}(\mathcal{E}_t, \mathcal{A}) = \mathcal{E}_{\text{reshape},\, t}, \qquad A_K \otimes \cdots \otimes A_v = A_V.$$

Example 1. Let $Y_t \in \mathbb{R}^{d_1 \times d_2}$ be a matrix-valued observation at $t \in [T]$. For the set-up, we can only specify $\mathcal{A} = \{1, 2\}$ (which is the only nontrivial case here, as discussed in § 3.1). The hypothesis test (6) is simplified as follows, with $\mathcal{A}$ reflected by the vectorization:

$$H_0: \; Y_t = A_1 F_t A_2^{\mathrm{T}} + E_t, \qquad H_1: \text{vec}(Y_t) = A_V \text{vec}(F_t) + \text{vec}(E_t) \text{ with } A_V \notin \mathcal{K}_{d_1 \times d_2}.$$

3.3. *Constructing the test statistic*

Despite the obscure $\mathcal{K}_{d_v \times \cdots \times d_K}$ in (6), we may resort to the Tucker TFM in (7) under H_0. To construct the test, we first obtain estimators for the standardized loading matrices in (5). For $j \in [v-1]$, $\widetilde{Q}_j$ is defined as the eigenvector matrix corresponding to the r_j largest eigenvalues of

$$\frac{1}{T} \sum_{t=1}^{T} \text{reshape}(\mathcal{Y}_t, \mathcal{A})_{(j)} \, \text{reshape}(\mathcal{Y}_t, \mathcal{A})_{(j)}^{\mathrm{T}},$$

where reshape$(\mathcal{Y}_t, \mathcal{A})_{(j)}$ is the mode-j unfolding matrix of reshape$(\mathcal{Y}_t, \mathcal{A})$; see § 2.1. Similarly, $\widetilde{Q}_V$ is the eigenvector matrix corresponding to the r_V largest eigenvalues of

$$\frac{1}{T} \sum_{t=1}^{T} \text{reshape}(\mathcal{Y}_t, \mathcal{A})_{(v)} \, \text{reshape}(\mathcal{Y}_t, \mathcal{A})_{(v)}^{\mathrm{T}}.$$

Then the estimators for $\mathcal{C}_{\text{reshape},\,t}$ and $\mathcal{E}_{\text{reshape},\,t}$ are respectively defined by

$$\widetilde{\mathcal{C}}_{\text{reshape},\,t} = \text{reshape}(\mathcal{Y}_t, \mathcal{A}) \times_{j=1}^{v-1} (\widetilde{Q}_j \widetilde{Q}_j^{\mathrm{T}}) \times_v (\widetilde{Q}_V \widetilde{Q}_V^{\mathrm{T}}),$$
$$\widetilde{\mathcal{E}}_{\text{reshape},\,t} = \text{reshape}(\mathcal{Y}_t, \mathcal{A}) - \widetilde{\mathcal{C}}_{\text{reshape},\,t}. \tag{8}$$

For (7), $\hat{Q}_j$ for $j \in [v-1]$ is defined as the eigenvector matrix corresponding to the r_j largest eigenvalues of $T^{-1}\sum_{t=1}^{T} Y_{t,(j)} Y_{t,(j)}^{\mathrm{T}}$. Next, let $\mathcal{R}$ be the set of all divisor combinations of r_V, i.e.,

$$\mathcal{R} = \left\{(\pi_1, \pi_2, \ldots, \pi_{K-v+1}) \text{ such that } \prod_{j=1}^{K-v+1} \pi_j = r_V \text{ and each } \pi_j \in \mathbb{Z}^+ \text{ with } \pi_j \leqslant d_{j+v-1}\right\}. \tag{9}$$

Let the mth element of $\mathcal{R}$ be $(\pi_{m,1}, \ldots, \pi_{m,K-v+1})$. For $i \in \{v, \ldots, K\}$, we define $\hat{Q}_{m,i}$ as the eigenvector matrix corresponding to the $\pi_{m,i-v+1}$ largest eigenvalues of $T^{-1}\sum_{t=1}^{T} Y_{t,(i)} Y_{t,(i)}^{\mathrm{T}}$. The common component and residual estimators are hence obtained as

$$\hat{\mathcal{C}}_{m,t} = \mathcal{Y}_t \times_{j=1}^{v-1} (\hat{Q}_j \hat{Q}_j^{\mathrm{T}}) \times_{i=v}^{K} (\hat{Q}_{m,i} \hat{Q}_{m,i}^{\mathrm{T}}),$$
$$\hat{\mathcal{E}}_{m,t} = \mathcal{Y}_t - \hat{\mathcal{C}}_{m,t}. \tag{10}$$

Let $\widetilde{\mathcal{E}}_t$ be the order-K tensor with the same dimension as $\mathcal{Y}_t$ such that $\text{reshape}(\widetilde{\mathcal{E}}_t, \mathcal{A}) = \widetilde{\mathcal{E}}_{\text{reshape},\,t}$. Define $k^* = \arg\min_{k \in [K]}\{d_k\}$ and denote the mode-k^* unfolding of $\widetilde{\mathcal{E}}_t$ and $\hat{\mathcal{E}}_{m,t}$ as $\widetilde{E}_{t,(k^*)}$ and $\hat{E}_{m,t,(k^*)}$, respectively. Theorem 2 (in § 4.2 below) tells us that there exists $m \in [|\mathcal{R}|]$ such that, for each $t \in [T]$, $j \in [d/d_{k^*}]$, both series defined by

$$x_{j,t} = \frac{1}{d_{k^*}} \sum_{i=1}^{d_{k^*}} \widetilde{E}^2_{t,(k^*),ij} \qquad y_{m,j,t} = \frac{1}{d_{k^*}} \sum_{i=1}^{d_{k^*}} \hat{E}^2_{m,t,(k^*),ij},$$

are asymptotically distributed the same under H_0, and $x_{j,t}$ in particular is distributed the same under either H_0 or H_1. Let $\mathbb{P}_{x,j}$ and $\mathbb{P}_{y,m,j}$, respectively, denote the empirical probability measures induced by the empirical cumulative distribution functions for $\{x_{j,t}\}_{t \in [T]}$ and $\{y_{m,j,t}\}_{t \in [T]}$:

$$\mathbb{F}_{x,j}(c) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{x_{j,t} \leqslant c\}, \qquad \mathbb{F}_{y,m,j}(c) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{y_{m,j,t} \leqslant c\}. \tag{11}$$

Let $\hat{q}_{x,j}(\alpha) = \inf\{c \mid \mathbb{F}_{x,j}(c) \geqslant 1-\alpha\}$. The intuition here is that if H_0 is satisfied then, over different $j \in [d/d_{k^*}]$, the cumulative distribution functions $\mathbb{F}_{x,j}(\cdot)$ and $\mathbb{F}_{y,m,j}(\cdot)$ should be similar. However, if H_1 is true then we expect the residuals in $\hat{E}_{m,t,(k^*)}$ to be inflated, so that $\mathbb{P}_{y,m,j}\{y_{m,j,t} \geqslant \hat{q}_{x,j}(\alpha)\}$ is expected to be larger than α; see the theoretical statement on this in Theorem 3 below. To incorporate this across different $j \in [d/d_{k^*}]$, we compare the

5% quantile of $T^{-1}\sum_{t=1}^{T}\mathbb{1}\{y_{m,j,t} \geqslant \hat{q}_{x,j}(\alpha)\}$ over $j \in [d/d_{k^*}]$ to α, and expect it to be larger than α under H_1.

Since, with the wrong number of factors, a particular $m \in [|\mathcal{R}|]$ will in general inflate the residuals $y_{m,j,t}$ further, in practice, to be on the conservative side, we reject H_0 if

$$\min_{m \in [|\mathcal{R}|]} \left\{ 5\% \text{ quantile of } \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\{y_{m,j,t} \geqslant \hat{q}_{x,j}(\alpha)\} \text{ over } j \in [d/d_{k^*}] \right\} > \alpha, \qquad (12)$$

noting that exactly one element in $\mathcal{R}$ represents the true number of factors on the modes with indices in $\mathcal{A}$. We also point out that there are other possible ways to aggregate the information from each j, but (12) empirically works well and circumvents possible issues such as heavy-tailed noise, underestimation or overestimation on the number of factors and insufficient data dimensions. See § 5.1 below.

Remark 1. It is possible to perform the test directly using the number of factors for modes in $\mathcal{A}$, i.e., $\mathcal{R}$ only contains the (either specified or estimated) number of factors $r_j, j = v, \ldots, K$, in (7). This is guaranteed by the Tucker TFM under H_0 in (6). However, in practice, we usually need to estimate the number of factors that would be invalid under H_1 in (6). This leads to unstable estimated ranks and hence an unstable test statistic, which we address by introducing $\mathcal{R}$ in (9).

4. ASSUMPTIONS AND THEORETICAL RESULTS

4.1. *Assumptions*

This subsection presents all the assumptions for testing H_0 against H_1 in (6). Another version (differing only in notation) of Assumptions 1 and 2 for the identification of (3) and (4) is included in the Supplementary Material.

Assumption 1. For each $j \in [v-1]$, we assume that A_j in (5) is of full rank and, as $d_j \to \infty$,

$$Z_j^{-1/2} A_j^{\mathrm{T}} A_j Z_j^{-1/2} \to \Sigma_{A,j}, \qquad (13)$$

where $\Sigma_{A,j}$ is positive definite with all eigenvalues bounded away from 0 and infinity, and Z_j is a diagonal matrix with $(Z_j)_{hh} \asymp d_j^{\delta_{j,h}}$ for $h \in [r_j]$ and the ordered factor strengths $1/2 < \delta_{j,r_j} \leqslant \cdots \leqslant \delta_{j,1} \leqslant 1$. We assume that A_V also has the above form with Z_V and $\Sigma_{A,V}$ in place of Z_j and $\Sigma_{A,j}$, respectively, except that only the maximum and minimum factor strengths are ordered, i.e., $1/2 < \delta_{V,r_V} \leqslant \delta_{V,h} \leqslant \delta_{V,1} \leqslant 1$ for any $h \in [r_V]$.

Assumption 2. With $\mathcal{A} = \{v, \ldots, K\}$, we assume that, for each $i \in \mathcal{A}$, A_i in (7) is of full rank and, as $d_i \to \infty$,

$$Z_i^{-1/2} A_i^{\mathrm{T}} A_i Z_i^{-1/2} \to \Sigma_{A,i},$$

where $\Sigma_{A,i}$ is positive definite with all eigenvalues bounded away from 0 and infinity, and Z_i is a diagonal matrix with $(Z_i)_{hh} \asymp d_i^{\delta_{i,h}}$ for $h \in [r_i]$ and the ordered factor strengths $1/2 < \delta_{i,r_i} \leqslant \cdots \leqslant \delta_{i,1} \leqslant 1$.

Assumption 3 (TIME SERIES IN $\mathcal{F}_{\text{reshape},t}$). There exists $\mathcal{X}_{\text{reshape},f,t}$ of the same dimension as $\mathcal{F}_{\text{reshape},t}$ such that $\mathcal{F}_{\text{reshape},t} = \sum_{w \geqslant 0} a_{f,w} \mathcal{X}_{\text{reshape},f,t-w}$. The time series $\{\mathcal{X}_{\text{reshape},f,t}\}$

has independent and identically distributed elements with mean 0, variance 1 and uniformly bounded fourth-order moments. The coefficients $a_{f,w}$ satisfy $\sum_{w\geqslant 0} a_{f,w}^2 = 1$ and $\sum_{w\geqslant 0} |a_{f,w}| \leqslant c$ for some constant c.

Assumption 4 (*Decomposition of* $\mathcal{E}_t$). The noise $\mathcal{E}_t$ (such that $\mathcal{E}_{\text{reshape},t} = \text{reshape}(\mathcal{E}_t, \mathcal{A})$) can be decomposed as

$$\mathcal{E}_t = \mathcal{F}_{e,t} \times_1 A_{e,1} \underset{2}{\times} \cdots \underset{K}{\times} A_{e,K} + \Sigma_\epsilon * \epsilon_t, \tag{14}$$

where order-K tensors $\mathcal{F}_{e,t} \in \mathbb{R}^{r_{e,1}\times\cdots\times r_{e,K}}$ and $\epsilon_t \in \mathbb{R}^{d_1\times\cdots\times d_K}$ contain independent mean-zero elements with unit variance, with the two time series $\{\epsilon_t\}$ and $\{\mathcal{F}_{e,t}\}$ being independent. The order-K tensor Σ_ϵ contains the standard deviations of the corresponding elements in ϵ_t, and has elements uniformly bounded. Moreover, for each $k \in [K]$, $A_{e,k} \in \mathbb{R}^{d_k \times r_{e,k}}$ is approximately sparse such that $\|A_{e,k}\|_1 = O(1)$.

Assumption 5 (*Time series in* $\mathcal{E}_t$). There exist $\mathcal{X}_{e,t}$ of the same dimension as $\mathcal{F}_{e,t}$ and $\mathcal{X}_{\epsilon,t}$ of the same dimension as ϵ_t such that $\mathcal{F}_{e,t} = \sum_{q\geqslant 0} a_{e,q}\mathcal{X}_{e,t-q}$ and $\epsilon_t = \sum_{q\geqslant 0} a_{\epsilon,q}\mathcal{X}_{\epsilon,t-q}$, with $\{\mathcal{X}_{e,t}\}$ and $\{\mathcal{X}_{\epsilon,t}\}$ independent of each other. Moreover, $\{\mathcal{X}_{e,t}\}$ has independent elements while $\{\mathcal{X}_{\epsilon,t}\}$ has independent and identically distributed elements, and all elements are mean zero with unit variance and uniformly bounded fourth-order moments. Both $\{\mathcal{X}_{e,t}\}$ and $\{\mathcal{X}_{\epsilon,t}\}$ are independent of $\{\mathcal{X}_{\text{reshape},f,t}\}$ from Assumption 3. The coefficients $a_{e,q}$ and $a_{\epsilon,t}$ are such that $\sum_{q\geqslant 0} a_{e,q}^2 = \sum_{q\geqslant 0} a_{\epsilon,q}^2 = 1$ and $\sum_{q\geqslant 0} |a_{e,q}|$, $\sum_{q\geqslant 0} |a_{\epsilon,q}| \leqslant c$ for some constant c.

Assumption 6 (*Rate assumptions*). With $g_s = \prod_{k=1}^K d_k^{\delta_{k,1}}$ and $\gamma_s = d_V^{\delta_{V,1}} \prod_{j=1}^{v-1} d_j^{\delta_{j,1}}$, assume that all of the following terms are $o(1)$:

$$dg_s^{-2}T^{-1}d_k^{2(\delta_{k,1}-\delta_{k,r_k})+1}, \qquad dg_s^{-1}d_k^{\delta_{k,1}-\delta_{k,r_k}-1/2},$$
$$d\gamma_s^{-2}T^{-1}d_V^{2(\delta_{V,1}-\delta_{V,r_V})+1}, \qquad d\gamma_s^{-1}d_V^{\delta_{V,1}-\delta_{V,r_V}-1/2}.$$

Assumption 7 (*Further rate assumptions*). With the definitions $g_w = \prod_{k=1}^K d_k^{\delta_{k,r_k}}$ and $\gamma_w = d_V^{\delta_{V,r_V}} \prod_{j=1}^{v-1} d_j^{\delta_{j,r_j}}$, we assume that all of the following terms are $o(d_{k^*}^{-1})$:

$$\max_{k\in[K]} \left\{ d_k^{2(\delta_{k,1}-\delta_{k,r_k})} \left(\frac{1}{Td_{\text{-}k}d_k^{1-\delta_{k,1}}} + \frac{1}{d_k^{1+\delta_{k,r_k}}} \right) \frac{d^2}{g_s g_w} \right\},$$
$$\max_{j\in[v-1]} \left\{ d_j^{2(\delta_{j,1}-\delta_{j,r_j})} \left(\frac{1}{Td_{\text{-}k}d_j^{1-\delta_{j,1}}} + \frac{1}{d_j^{1+\delta_{j,r_j}}} \right) \frac{d^2}{\gamma_s \gamma_w} \right\},$$
$$d_V^{2(\delta_{V,1}-\delta_{V,r_V})} \left(\frac{1}{Tdd_V^{-\delta_{V,1}}} + \frac{1}{d_V^{1+\delta_{V,r_V}}} \right) \frac{d^2}{\gamma_s \gamma_w}, \qquad \frac{d}{\gamma_w^2}, \qquad \frac{d}{g_w^2}.$$

With Assumption 1, the standardized loading matrix defined by $Q_j = A_j Z_j^{-1/2}$ satisfies $Q_j^{\mathrm{T}} Q_j \to \Sigma_{A,j}$ for $j \in [v-1]$, and $Q_V = A_V Z_V^{-1/2}$ satisfies $Q_V^{\mathrm{T}} Q_V \to \Sigma_{A,V}$. A similar implication holds for Assumption 2, except that Assumption 2 is only valid under the null. Hence, with Assumption 2, Z_V and $\Sigma_{A,V}$ in Assumption 1 satisfy

$$Z_V = Z_K \otimes \cdots \otimes Z_v, \qquad \Sigma_{A,V} = \Sigma_{A,K} \otimes \cdots \otimes \Sigma_{A,v}. \tag{15}$$

The factor strength requirement for A_V in Assumption 1 is satisfied by Assumption 2, since, from (15),

$$(Z_V)_{r_V r_V} \asymp \prod_{i=v}^{K} d_i^{\delta_{i,r_i}} \geqslant d_V^{\delta_{\min}} > d_V^{1/2},$$

where $\delta_{\min} = \min_{i=v,\ldots,K}\{\delta_{i,r_i}\}$. Assumption 1 characterizes the loading matrix behaviour generally for (6), and the additional Assumption 2 is specific for the null. Both assumptions allow for weak factors that are common features in the literature (Lam & Yao, 2012; Onatski, 2012; Cen & Lam, 2025). Under the assumption that all factors are pervasive, (13) can be interpreted as $d_j^{-1} A_j^{\mathrm{T}} A_j \to \Sigma_{A,j}$, aligning with Assumption 3 of Chen & Fan (2023) for matrix time series.

Assumption 3 assumes that $\mathcal{F}_{\text{reshape},t}$ is a general linear process with weak serial dependence. Theorem 1 ensures that the core factor in (7) (under H_0) retains the structure of Assumption 3 such that

$$\mathcal{F}_t = \sum_{w \geqslant 0} a_{f,w} \mathcal{X}_{f,t-w}$$

with $\text{reshape}(\mathcal{X}_{f,t}, \mathcal{A}) = \mathcal{X}_{\text{reshape},f,t}$. For each $k \in [K]$, as $T \to \infty$, we have the convergence in probability that

$$\frac{1}{T} \sum_{t=1}^{T} \text{mat}_k(\mathcal{F}_t)\, \text{mat}_k(\mathcal{F}_t)^{\mathrm{T}} \to r_{\text{-}k}\, I_{r_k}, \tag{16}$$

where I_{r_k} is the $r_k \times r_k$ identity matrix, which is direct from Proposition 1.3 in the supplement of Cen & Lam (2025). In comparison, Barigozzi et al. (2025) assumed the form of (16) with $r_{\text{-}k}\, I_{r_k}$ replaced by a positive definite matrix. This does not imply that Assumption 3 is particularly stronger as our factor loading matrices have already incorporated all potential correlations through Assumptions 1 and 2.

Assumptions 4 and 5 depict a general noise time series on the factor models (5) and (7). The noise tensor $\mathcal{E}_t$ is allowed to be weakly dependent across modes and time, regardless of the existence of the Kronecker product structure. From (14), we have

$$\begin{aligned}\text{reshape}(\mathcal{E}_t, \mathcal{A}) &= \text{reshape}(\mathcal{F}_{e,t}, \mathcal{A}) \mathop{\times}_{j=1}^{v-1} A_{e,j} \times_v (A_K \otimes \cdots \otimes A_v) \\ &\quad + \text{reshape}(\Sigma_\epsilon, \mathcal{A}) * \text{reshape}(\epsilon_t, \mathcal{A}),\end{aligned}$$

so that the structure of Assumptions 4 and 5 are preserved by $\text{reshape}(\mathcal{E}_t, \mathcal{A})$. Assumption 6 details the rate assumptions on factor strengths and is hence satisfied automatically when all factors are pervasive. Assumption 7 also concerns factor strength and holds when all factors are pervasive if $v > 1$; for $v = 1$, Assumption 7 holds when $\min_{k \in [K]} d_k = o(T)$ in addition to pervasive factors.

Remark 2. When $v = 1$ and all factors are pervasive, Assumption 7 requires $d_{k*} = \min_{k \in [K]} d_k = o(T)$, which seems restrictive. This condition was imposed to ensure asymptotic normality when aggregating d_{k*} estimated residuals in $x_{j,t}$ and $y_{m,j,t}$ in § 3.3. However, from the proof of Theorem 2 in the Supplementary Material, it is feasible to aggregate d_{k*}^{β} for any $0 < \beta < 2$ such that $d_{k*}^{\beta} = o(T)$. Therefore, Assumption 7 is in fact arguably as mild as Assumption B5 of He et al. (2023). We do not pursue such an aggregation scheme here to keep the practical procedure as simple as possible, but we illustrate this with numerical results in the Supplementary Material.

4.2. *Main results and practical test design*

We first present below the results for our residual estimators in (8) and (10), which inspired the testing procedure in § 3.3. Following Theorem 2, the theoretical guarantee of the test is also provided.

THEOREM 2. *Let Assumptions 1–7 hold. Using the notation in § 3.3, under H_0, there exists $m \in [|\mathcal{R}|]$ such that, for each $t \in [T]$, $j \in [d/d_k^*]$,*

$$\frac{\sum_{i=1}^{d_{k*}} (\hat{E}^2_{m,t,(k^*),ij} - \Sigma^2_{\epsilon,(k^*),ij})}{\{\sum_{i=1}^{d_{k*}} \mathrm{var}(\epsilon^2_{t,(k^*),ij})\Sigma^4_{\epsilon,(k^*),ij}\}^{1/2}}, \quad \frac{\sum_{i=1}^{d_{k*}} (\widetilde{E}^2_{t,(k^*),ij} - \Sigma^2_{\epsilon,(k^*),ij})}{\{\sum_{i=1}^{d_{k*}} \mathrm{var}(\epsilon^2_{t,(k^*),ij})\Sigma^4_{\epsilon,(k^*),ij}\}^{1/2}} \to Z_{j,t},$$

in probability, where $Z_{j,t}$ is asymptotically standard normal and $Z_{h,t}$ is independent of $Z_{\ell,t}$ for $h \neq \ell$. Under H_1, the asymptotic result for $\widetilde{E}_{t,(k^*),ij}$ above still holds.

THEOREM 3. *Let all the assumptions in Theorem 2 hold. In addition, each element in the time series $\{\mathcal{X}_{\text{reshape},f,t}\}$, $\{\mathcal{X}_{e,t}\}$ and $\{\mathcal{X}_{\epsilon,t}\}$ has sub-Gaussian tail. Using the notation in § 3.3, for any $j \in [d/d_{k^*}]$ under H_0, there exists $m \in [|\mathcal{R}|]$ such that, as* $\min\{T, d_1, \ldots, d_K\} \to \infty$,

$$\mathbb{P}_{y,m,j}\{y_{m,j,t} > \hat{q}_{x,j}(\alpha)\} \leqslant \alpha + O_P(\rho),$$

where

$$\rho = \left[\max_{k \in [K]} \left\{ d_k^{\delta_{k,1} - \delta_{k,r_k}} \left(\frac{1}{(Td_{\text{-}k} d_k^{1-\delta_{k,1}})^{1/2}} + \frac{1}{d_k^{(1+\delta_{k,r_k})/2}} \right) \frac{d}{(g_s g_w)^{1/2}} \right\} + \frac{d^{1/2}}{g_w} \right]$$
$$\times \log^2(T) \left(\prod_{k=1}^{K} \log^2(d_k) \right) d_{k^*}^{1/2}$$
$$+ \log^2(T) \log(d_V) \left(\prod_{k=1}^{v-1} \log(d_k) \right) \left(\prod_{k=1}^{K} \log^2(d_k) \right) d_{k^*}^{1/2}$$

$$\times \left[\max_{j\in[v-1]} \left\{ d_j^{\delta_{j,1}-\delta_{j,r_j}} \left(\frac{1}{(Td_{\text{-}k}d_j^{1-\delta_{j,1}})^{1/2}} + \frac{1}{d_j^{(1+\delta_{j,r_j})/2}} \right) \frac{d}{(\gamma_s\gamma_w)^{1/2}} \right\} + d_V^{\delta_{V,1}-\delta_{V,r_V}} \left(\frac{1}{(Tdd_V^{-\delta_{V,1}})^{1/2}} + \frac{1}{d_V^{(1+\delta_{V,r_V})/2}} \right) \frac{d}{(\gamma_s\gamma_w)^{1/2}} + \frac{d^{1/2}}{\gamma_w} \right].$$

Theorem 3 suggests that if some factors are weaker then the rate in the probability statement above will be inflated. When all factors are pervasive, define $d_{\max} = \max_{k\in[K]}\{d_k\}$, and we may simplify ρ as

$$\rho = \left\{ \frac{1}{d_{k^*}^{1/2}} + \left(\frac{d_{k^*}d_{\max}}{Td}\right)^{1/2} + \left(\frac{d_{k^*}d_V}{Td}\right)^{1/2} \log(d_V) \left(\prod_{k=1}^{v-1} \log(d_k) \right) \right\} \times \log^2(T) \left(\prod_{k=1}^{K} \log^2(d_k) \right).$$

Hence, $\rho = o(1)$ as long as $T, d_1, \ldots, d_K$ are of the same order, but it appears that, when $d_V = d$, i.e., $\mathcal{A} = [K]$, the current test requires $d_{k^*} \log(d) \log^2(T) \prod_{k=1}^{K} \log^2(d_k) = o(T)$. However, this can be circumvented, as explained in Remark 2. Theorem 3 presents the grounds for our construction of the test statistic in (12). For related explanations, see the discussions immediately after (11), and before Remark 1.

The set-up of problem (6) specifies set $\mathcal{A}$, which is only needed in H_1 due to (7) under H_0. It is straightforward to specify $\mathcal{A}$ for a series of matrix-valued observations (i.e., an order-2 tensor); see Example 1. However, for a general order-K tensor with $K \geqslant 3$, $\mathcal{A}$ might be misspecified without any prior knowledge. To resolve this, we now present the second theorem on tensor reshaping, for both the Tucker TFM and CP TFM for completeness.

THEOREM 4 (TENSOR RESHAPING II). *Consider a tensor time series $\{\mathcal{Y}_t\}$ and a set of mode indices $\mathcal{A}$. With Definition 2, the time series $\{\text{reshape}(\mathcal{Y}_t, \mathcal{A})\}$ following a Tucker TFM has a Kronecker product structure if and only if $\{\mathcal{Y}_t\}$ has either a Kronecker product structure or no Kronecker product structure along a subset of $\mathcal{A}$.*

The above remains true upon replacing a Tucker TFM by a CP TFM and the Kronecker product structure by the Khatri–Rao product structure.

Suppose now that $\{\mathcal{Y}_t\}$ has no Kronecker product structure along some $\mathcal{A}^*$. Theorem 4 tells us that testing the Kronecker product structure of the reshaped series $\{\text{reshape}(\mathcal{Y}_t, \mathcal{A})\}$ effectively tests if $\mathcal{A}^* \subseteq \mathcal{A}$. In light of this, a testing design is feasible when $\mathcal{A}$ is unspecified, with a minimal assumption that $\text{reshape}(\mathcal{Y}_t, [K]) = \text{vec}(\mathcal{Y}_t)$ has a factor structure, i.e., the vectorized $\mathcal{Y}_t$ follows a standard vector factor model. For illustration, consider $\text{reshape}(\mathcal{Y}_t, [K]\setminus\{1\}) = \text{reshape}(\mathcal{Y}_t, \{2, \ldots, K\})$, which is an order-2 tensor. Using the property of $\text{reshape}(\cdot,\cdot)$ in Appendix A in the Supplementary Material, we have

$$\text{reshape}\{\text{reshape}(\mathcal{Y}_t, \{2, \ldots, K\}), \{1, 2\}\} = \text{reshape}(\mathcal{Y}_t, \{1, \ldots, K\}) = \text{vec}(\mathcal{Y}_t).$$

According to Definition 2, $\{\text{reshape}(\mathcal{Y}_t, \{2, \ldots, K\})\}$ follows a Tucker TFM along $\{1, 2\}$. This is always correctly specified since $\{\text{vec}(\mathcal{Y}_t)\}$ follows a factor model (which also implies that $\mathcal{A}^* \subseteq [K]$). By Theorem 4, $\text{reshape}(\mathcal{Y}_t, \{2, \ldots, K\})$ has no Kronecker product structure if and only if $1 \in \mathcal{A}^*$. Hence, in testing (6) with $\mathcal{Y}_t$ replaced by $\{\text{reshape}(\mathcal{Y}_t, \{2, \ldots, K\})\}$ and $\mathcal{A} = \{1, 2\}$, rejection of the null implies that $1 \in \mathcal{A}^*$.

By the fact that $\{\mathrm{vec}(\mathcal{Y}_t)\}$ with any permutation on $\mathrm{vec}(\mathcal{Y}_t)$ also follows a factor model, the above scheme is in fact valid on $\mathrm{reshape}(\mathcal{Y}_t, [K] \setminus \{k\})$ for any $k \in [K]$. Eventually, $\mathcal{A}^*$ can be identified, and the above procedure is summarized in the following algorithm.

Algorithm 1. A practical testing algorithm, given an order-K tensor time series $\{\mathcal{Y}_t\}$ with $K \geqslant 2$ and $vec(\mathcal{Y}_t)$ following a factor model with r_{vec} known factors.

Set $\hat{\mathcal{A}}^* = \emptyset$
For $k = 1$ to $k = K$
  Define a test as (6) with $\{\mathcal{Y}_t\}$ replaced by $\{\textsc{reshape}(\mathcal{Y}_t, [K] \setminus \{k\})\}$ and $\mathcal{A}$ by $\{1, 2\}$
  Test the problem in the previous line using the testing procedure in § 3.3 with r_V replaced by r_{vec}
  If the null is rejected
    $\hat{\mathcal{A}}^* \leftarrow \hat{\mathcal{A}}^* \cup \{k\}$
Output $\hat{\mathcal{A}}^*$

With the output from the algorithm, we conclude that $\{\mathcal{Y}_t\}$ has no Kronecker product structure along $\hat{\mathcal{A}}^*$. In practice, $\hat{\mathcal{A}}^*$ being an empty set implies that $\{\mathcal{Y}_t\}$ has a Kronecker product structure.

5. Numerical studies

5.1. *Simulations*

In this subsection, we demonstrate the empirical performance of our test with respect to hypothesis (6) using Monte Carlo simulations. As discussed in § 3.1, the test is only nontrivial when the data order K is at least 2. We therefore consider $K = 2$ to $K = 4$.

The data-generating processes adapt Assumptions 3, 4 and 5. Specifically, we set the number of factors as $r_k = 2$ for any $k \in [K]$, and first generate $\mathcal{F}_t$ in (7) with each element being an independent standardized autoregressive model of order 2 with autoregressive coefficients 0.7 and -0.3. The elements in $\mathcal{F}_{e,t}$ and ϵ_t are generated similarly, but their autoregressive coefficients are $(-0.5, 0.5)$ and $(0.4, 0.4)$, respectively. The standard deviation of each element in ϵ_t is generated by independent and identically distributed $|\mathcal{N}(0, 1)|$. Unless specified otherwise, all innovation processes in constructing $\mathcal{F}_t$, $\mathcal{F}_{e,t}$ and ϵ_t are independent and identically distributed standard normal. For each $j \in [v-1]$, each factor loading matrix A_j is generated independently with $A_j = U_j B_j$, where each entry of $U_j \in \mathbb{R}^{d_j \times r_j}$ is independent and identically distributed $\mathcal{N}(0, 1)$, and $B_j \in \mathbb{R}^{r_j \times r_j}$ is diagonal with the hth diagonal entry being $d_j^{-\zeta_{j,h}}$, $0 \leqslant \zeta_{j,h} \leqslant 0.5$. Pervasive factors have $\zeta_{j,h} = 0$, while weak factors have $0 < \zeta_{j,h} \leqslant 0.5$. Each entry of $A_{e,j} \in \mathbb{R}^{d_j \times r_{e,j}}$ is independent and identically distributed $\mathcal{N}(0, 1)$, but with a 0.95 probability of being set exactly to 0. We set $r_{e,k} = 2$ for all $j \in [v-1]$ throughout all experiments. For any $\mathcal{A}$ (specified later), we obtain

$$\mathrm{reshape}(\mathcal{F}_t, \mathcal{A}) = \mathcal{F}_{\mathrm{reshape},\,t}, \qquad \mathrm{reshape}(\mathcal{E}_t, \mathcal{A}) = \mathcal{E}_{\mathrm{reshape},t}.$$

Lastly, similar to $\{A_j\}_{j \in [v-1]}$, we generate $\{A_v, \ldots, A_K\}$ and let $A_V = A_K \otimes \cdots \otimes A_v$ under H_0, or generate A_V directly under H_1. Whenever r_V is required, it is computed as $\prod_{j \in \mathcal{A}} r_j$, which has $2^{|\mathcal{A}|}$ combinations since we set two factors for each mode. According to (5) and

Table 1. *Test performance under H_0 for various settings*

		$K=2$				$K=3$				$K=4$			
		$d_k=15$		$d_k=30$		$d_k=15$		$d_k=30$		$d_k=10$		$d_k=15$	
	T	$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$
$\hat{\alpha}$	120	2	7.1	2	7.8	1.3	5.5	1.3	5.5	1.2	5.4	1.2	5.3
	360	1.1	5.7	1.2	5.9	1	5.1	1	5.2	1	5.1	1	5.1
	720	1.1	5.3	1.2	5.4	1	5.1	1	5.1	1	5	1	5.1
$\hat{p}$	120	97.4	83.6	99.6	86	100	100	100	100	100	100	100	100
	360	98.8	86.2	100	84.2	100	100	100	100	100	100	100	100
	720	99.4	91.6	100	92	100	100	100	100	100	100	100	100

For each setting, d_k is the same for all $k \in [K]$. Each cell is the average of $\hat{\alpha}$ or $\hat{p}$ computed under the corresponding setting over 500 runs. All values have been multiplied by 10^2.

(7), we then respectively construct $\text{reshape}(\mathcal{Y}_t, \mathcal{A})$ (and hence the corresponding $\mathcal{Y}_t$) or $\mathcal{Y}_t$ directly.

We consider a series of performance indicators and each simulation setting is repeated 500 times. Using the notation in § 3.3, with $\alpha \in \{0.01, 0.05\}$, we calculate

$$\hat{\alpha} = \min_{m \in [|\mathcal{R}|]} \left\{ \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\{y_{m,1,t} \geqslant \hat{q}_{x,1}(\alpha)\} \right\}, \qquad \hat{p} = \mathbb{1}\{\hat{q}_\alpha \leqslant \alpha\},$$
$$\text{with} \quad \hat{q}_\alpha = \min_{m \in [|\mathcal{R}|]} \left\{ 5\% \text{ quantile of } \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\{y_{m,j,t} \geqslant \hat{q}_{x,j}(\alpha)\} \text{ over } j \in [d/d_{k^*}] \right\}, \tag{17}$$

where $\hat{\alpha}$ is the significance level under measure $\mathbb{P}_{y,m,1}$, taken as the minimum over $m \in [|\mathcal{R}|]$, and $\hat{p}$ is an indicator function of the decision rule (12) that leads to retaining H_0. Under H_0, we expect $\hat{\alpha}$ to be close to α and $\hat{p}$ to be 1 according to Theorem 3.

Consider first H_0 with $\mathcal{A}$ containing the last two modes of $\mathcal{Y}_t$, i.e., $\mathcal{A} = \{1, 2\}$ for $K = 2$, $\mathcal{A} = \{2, 3\}$ for $K = 3$ and $\mathcal{A} = \{3, 4\}$ for $K = 4$. We experiment on all pervasive factors. Table 1 presents the simulation results under various settings for $K = 2, 3, 4$, and all of them align well with Theorem 3. For $K = 3, 4$, we have $\hat{p} = 1$ throughout, and for $K = 2$, the proportion of repetitions with $\hat{p} = 1$ generally increases with dimensions and time. The results under H_1 are presented in Table 2, which confirm the power of our test. While larger dimensions generally improve the test performance, it is unsurprising from Table 2 that, under the same (T, d_k) setting, testing the Kronecker product structure along two modes on $\mathcal{Y}_t$ is more difficult for higher order $\mathcal{Y}_t$. This is reasonable since the testing problem (6) is genuinely harder when A_V plays a less significant role in a higher-order dataset. To demonstrate this, suppose that $K = 3$, $(T, d_1, d_2, d_3) = (360, 10, 15, 20)$, and that all factors are pervasive. We experiment through $\mathcal{A} = \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$. The results, reported in Table 3, indeed show that, when the tested loading matrix A_V has a larger size, the test has larger power in general. The setting with $\mathcal{A} = \{2, 3\}$ is an exception, suggesting a potential issue with unbalanced spatial dimensions.

In the following, we fix $K = 3$ and $\mathcal{A} = \{2, 3\}$ to investigate the robustness of our test. Consider settings I and II, each with four subsettings:

(Ia) $T = 180$, $d_1 = d_2 = d_3 = 15$ and all factors pervasive with $\zeta_{j,h} = 0$,
(Ib) same as (Ia), but one factor is weak with $\zeta_{j,1} = 0.1$,

Table 2. *Test performance under H_1 for various settings*

		$K=2$				$K=3$				$K=4$			
		$d_k=15$		$d_k=30$		$d_k=15$		$d_k=30$		$d_k=10$		$d_k=15$	
	T	$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$
$\hat{\alpha}$	120	83.9	89.8	92.8	95.6	67.4	74.2	79	83.4	58.3	65.5	64.9	71.2
	360	81.8	88.8	91.7	95.1	65.9	73.8	77.6	83.2	57.1	65.3	63.6	70.9
	720	81.7	88.5	91.8	95.1	65.2	73.1	78.7	83.7	55.9	64	62.9	70.1
$\hat{p}$	120	0	0	0	0	0	0	0	0	1.2	0.2	0	0
	360	0	0	0	0	0	0	0	0	0.2	0	0	0
	720	0	0	0	0	0	0	0	0	0.2	0	0	0

See Table 1 for an explanation of each cell. All values have been multiplied by 10^2.

Table 3. *Test performance over different $\mathcal{A}$ with dimension* $(T, d_1, d_2, d_3) = (360, 15, 20, 25)$

		$\mathcal{A}=\{1,2\}$		$\mathcal{A}=\{1,3\}$		$\mathcal{A}=\{2,3\}$		$\mathcal{A}=\{1,2,3\}$	
		$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$
H_0	$\hat{\alpha}$	1	5.1	1	5.2	1	5.2	1.4	6.3
	$\hat{p}$	100	100	100	100	100	100	100	95.6
H_1	$\hat{\alpha}$	70.2	77.5	70.5	77.9	67.3	74.8	92.7	95.9
	$\hat{p}$	0	0	0	0	0	0	0	0

See Table 1 for an explanation of each cell. The number of rows of A_V in (6) is 300, 375, 500, 7500 for $\mathcal{A} = \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}$, respectively. All values have been multiplied by 10^2.

(Ic) same as (Ia), but both factors are weak with $\zeta_{j,1} = \zeta_{j,2} = 0.1$,
(Id) same as (Ia), but all innovation processes in constructing $\mathcal{F}_t$, $\mathcal{F}_{e,t}$ and ϵ_t are independent and identically distributed t_3,
(IIa)–(IId) same as (Ia)–(Id), respectively, except that r_V is randomly specified from 2–6 with equal probability.

Setting (Ia) is our benchmark and all other settings feature some defects from weak factors, heavy-tailed noise or a misspecified number of factors. Table 4 reports the results for both H_0 and H_1. In contrast to (Ia), all other settings have lower test power to various extents. However, the size of the test is hardly affected by weak factors or heavy-tailed noise, as shown by the results for settings (Ib), (Ic) and (Id). Although misspecification of the number of factors is detrimental, our decision rule $\hat{p}$ still performs satisfactorily.

On the practical testing algorithm that does not require $\mathcal{A}$ to be specified, we consider settings III and IV with $K = 3$, each with three subsettings:

(IIIa) $T = 360$, $d_k = 10$, all factors are strong and the series has a Kronecker product structure,
(IIIb) same as (IIIa), but the data have no Kronecker product structure along $\{2, 3\}$,
(IIIc) same as (IIIa), but the data have no Kronecker product structure along $\{1, 2, 3\}$,
(IVa)–(IVc) same as (IIIa)–(IIIc), respectively, except that $T = 720$.

Table 5 verifies that our algorithm is able to test the Kronecker product structure of given data without prespecifying $\mathcal{A}$. The performance improves with more observations, and the level $\alpha = 0.01$ works particularly well.

Table 4. *Test performance over the subsettings of settings I and II*

		(Ia)		(Ib)		(Ic)		(Id)	
		$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$
H_0	$\hat{\alpha}$	0.8	5.4	0.8	5.3	0.8	5.3	0.8	5.3
	$\hat{p}$	100	100	100	100	100	100	100	100
H_1	$\hat{\alpha}$	69.1	76.5	59.3	68.4	44.1	55.3	51.9	69.3
	$\hat{p}$	0	0	1.4	0	3.4	0.4	7	0.2
		(IIa)		(IIb)		(IIc)		(IId)	
		$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$
H_0	$\hat{\alpha}$	5.4	11.3	3.5	9.1	2.1	7.4	3	9.2
	$\hat{p}$	97.2	93.2	98.8	96.6	99.8	99.6	99.6	96.4
H_1	$\hat{\alpha}$	55.3	65.2	50.4	62	37.8	50.9	42.4	59.6
	$\hat{p}$	3.4	0	1.8	0.2	3.6	0.6	12.8	0.4

See Table 1 for an explanation of each cell. All values have been multiplied by 10^2.

Table 5. *Performance of the practical testing algorithm over the subsettings of settings III and IV*

	(IIIa)		(IIIb)		(IIIc)		(IVa)		(IVb)		(IVc)	
	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$
Mode 1	0	2.4	3	20.2	100	100	0	0	0.4	22.8	100	100
Mode 2	0	3.6	100	100	100	100	0	0.4	100	100	100	100
Mode 3	0	3.4	99.8	100	100	100	0	0.2	100	100	100	100

Each cell is the fraction, multiplied by 10^2, of the corresponding mode identified over 500 runs for the corresponding subsettings.

5.2. *Real data analysis*

We apply our test on four real data examples, described as follows.

(i) *New York City taxi traffic*. The dataset includes all individual taxi rides operated by Yellow Taxi within Manhattan Island of New York City; see https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page for further details. It contains trip records for the period from 1 January 2018 to 31 December 2022. We focus on the pick-up and drop-off dates/times, and the pick-up and drop-off locations that are coded according to 69 predefined zones in the dataset. Moreover, each day is divided into 24 hourly periods to represent the pick-up and drop-off times, with the first hourly period spanning from 12 a.m. to 1 a.m. Hence, each day we have $\mathcal{Y}_t \in \mathbb{R}^{69\times 69\times 24}$, where $y_{i_1,i_2,i_3,t}$ is the number of trips from zone i_1 to zone i_2 and the pick-up time is within the i_3th hourly period on day t. We consider business days and nonbusiness days separately, so that we analyse two tensor time series. The business-day series and the nonbusiness-day series are 1260 and 566 days long, respectively.

(ii) *Fama-French portfolio returns.* This is a set of portfolio return data, where stocks are respectively categorized into ten levels based on market equity and the book-to-equity ratio, where the latter is defined as the book equity from the last fiscal year divided by the end-of-year market equity; both criteria use NYSE deciles as breakpoints at the end of June each year. See https://mba.tuck.dartmouth.edu/pages/

faculty/ken.french/Data_Library/det_100_port_sz.html for further details. The stocks in each of the 10×10 categories form exactly two portfolios: one being value weighted and the other equally weighted. That is, we study two sets of 10×10 portfolios along with their time series. We use monthly data from January 2010 to June 2021, and hence, for both value-weighted and equally weighted portfolios, we have each of our datasets is an order-2 tensor $\mathcal{X}_t \in \mathbb{R}^{10\times 10}$ for $t \in [138]$.

(iii) *Image recognition.* This is the testing set from the Modified National Institute of Standards and Technology dataset with 10 000 samples and has also been analysed by He et al. (2023). Each image $\mathcal{Y}_t$ contains 28×28 pixels and is grey scale to represent handwritten digit numbers 0–9.

(iv) *Macroeconomic indices from the Organization for Economic Cooperation and Development database.* We use the same dataset, kindly provided by the authors, as He et al. (2023), which was also analysed by Yu et al. (2022a). The data include 10 macroeconomic indicators for eight countries (the United States, the United Kingdom, Canada, France, Germany, Norway, Australia and New Zealand) over 130 quarters from 1988-Q1 to 2020-Q2. Each series is transformed and standardized according to Yu et al. (2022a).

The two taxi series are order-3 tensor time series with the first two modes representing locations and the third representing time; hence, for interpretability, we only test their Kronecker product (and Khatri–Rao product) structure along $\mathcal{A} = \{1, 2\}$, i.e., we speculate that there is a merged location factor instead of pick-up and drop-off factors along modes 1 and 2, respectively. On the other hand, all the remaining series are order-2 tensor time series, so we test along $\mathcal{A} = \{1, 2\}$, as explained in § 3.1. Furthermore, we remove the market effect via the capital asset pricing model (CAPM) as

$$\mathrm{vec}(\mathcal{X}_t) = \mathrm{vec}(\bar{\mathcal{X}}) + (r_t - \bar{r})\beta + \mathrm{vec}(\mathcal{Y}_t),$$

where $\mathrm{vec}(\mathcal{X}_t) \in \mathbb{R}^{100}$ is the vectorized returns at time t, $\mathrm{vec}(\bar{\mathcal{X}})$ is the sample mean of $\mathrm{vec}(\mathcal{X}_t)$, β is the coefficient vector, r_t is the return of the NYSE composite index at time t, $\bar{r}$ is the sample mean of r_t and $\mathrm{vec}(\mathcal{Y}_t)$ is the CAPM residual. The least-squares solution is

$$\hat{\beta} = \frac{\sum_{t=1}^{138}(r_t - \bar{r})\{\mathrm{vec}(\mathcal{X}_t) - \mathrm{vec}(\bar{\mathcal{X}})\}}{\sum_{t=1}^{138}(r_t - \bar{r})^2},$$

so that the estimated residual series $\{\hat{\mathcal{Y}}_t\}_{t\in[138]}$ with $\hat{\mathcal{Y}}_t \in \mathbb{R}^{10\times 10}$ is constructed as $\{\mathrm{vec}(\mathcal{X}_t) - \mathrm{vec}(\bar{\mathcal{X}}) - (r_t - \bar{r})\hat{\beta}\}_{t\in[138]}$.

Hence, we study eight time series in total: business-day taxi series, nonbusiness-day taxi series, value-weighted portfolio series, equally weighted portfolio series, value-weighted residual series, equally weighted residual series, image recognition series and macroeconomic index series. For each series, we perform the test described in § 3.3. We also added the results for testing the Khatri–Rao product structure for a CP TFM in Table 6 for these eight time series. To estimate the number of factors for a Tucker TFM, we use the approaches BCorTh by Chen & Lam (2024), iTIP-ER by Han et al. (2022) and RTFA-ER by He et al. (2022) directly on each time series due to their large dimensions. Each mode of the first six series has one or two estimated factors. Since the test results are similar for those rank settings, we present the results with two factors on each mode and, hence, $\hat{r}_V = 4$. Hence, we

Table 6. *Test results for the studied series*

	Tucker TFM				CP TFM				Tests in He et al. (2023)
	$\hat{\alpha}$		$\hat{q}_\alpha$		$\hat{\alpha}$		$\hat{q}_\alpha$		
	1%	5%	1%	5%	1%	5%	1%	5%	
Business-day taxi	1.8	9.3	0.2	0.3	1.8	9.3	0.2	0.4	–
Nonbusiness-day taxi	1.8	9.5	0.4	1.1	1.8	9.5	0.5	1.9	–
Value-weighted portfolio	3.6	8.7	**1.1**	**5.3**	5.8	12.3	**1.4**	**6.8**	Not rejected
Equally weighted portfolio	3.6	5.1	**1.8**	3.9	4.3	7.2	**1.8**	**5.8**	Not rejected
Value-weighted residual	2.2	6.5	**1.1**	4.7	2.2	11.6	**1.4**	**5.4**	Not rejected
Equally weighted residual	1.4	5.1	**1.1**	4.7	2.2	5.1	**1.1**	4.7	Not rejected
Image recognition	0.0	0.0	0.2	0.4	0.2	0.6	1.0	2.8	Not rejected
Macroeconomic indices	5.5	10.2	**2.3**	**6.2**	5.5	10.2	**2.3**	**6.2**	Not rejected

Columns 1–4 report the results for our hypothesis of interest (6), i.e., under the Tucker TFM, with $\mathcal{A} = \{1, 2\}$. Columns 5–8 report similar results, but for testing the Khatri–Rao product structure under the CP TFM. We denote $\hat{q}_\alpha$ in **bold** if $\hat{q}_\alpha > \alpha$. The last column reports the test results according to He et al. (2023) for H_0 versus either $H_{1,\text{row}}$ or $H_{1,\text{col}}$, with significance levels 1% and 5% both experimented. All values have been multiplied by 10^2.

use two factors in the testing of the Khatri–Rao product for a CP TFM in the first six time series in Table 6. Following Yu et al. (2022a) and He et al. (2023) for a direct comparison, we use $(\hat{r}_1, \hat{r}_2) = (4, 5)$ for the image recognition series, whereas, for the economic indicator series, we experiment with all combinations of $\hat{r}_1 = 1, 2$ and $\hat{r}_2 = 4, 5$, and present the result for $(\hat{r}_1, \hat{r}_2) = (2, 4)$ as the conclusions are similar.

In addition, we also conduct the hypotheses tests of He et al. (2023) on our matrix time series datasets. To explain their hypotheses, for a matrix time series $\{Y_t\}$ with $Y_t \in \mathbb{R}^{d_1 \times d_2}$, under the null, (7) for $K = 2$, we have

$$H_0\colon\ Y_t = A_1 F_t A_2^{\mathrm{T}} + E_t,$$

where $F_t \in \mathbb{R}^{r_1 \times r_2}$. However, under their two alternatives, we test

$$H_{1,\text{row}}\colon r_2 = 0, \qquad H_{1,\text{col}}\colon r_1 = 0,$$

where, according to He et al. (2023), $r_1 > 0, r_2 = 0$ (respectively $r_2 > 0, r_1 = 0$) denotes a one-way factor model along the row dimension, so that $Y_t = A_1 F_{1,t} + E_t$ with $F_{1,t} \in \mathbb{R}^{r_1 \times d_2}$ (respectively the column dimension, so that $Y_t = F_{2,t} A_2^{\mathrm{T}} + E_t$ with $F_{2,t} \in \mathbb{R}^{d_1 \times r_2}$), and $r_1 = r_2 = 0$ denotes the absence of any factor structure, so that $Y_t = E_t$. All hyperparameter set-ups in Table 8 and 9 of He et al. (2023) are experimented on and all conclusions are the same.

Table 6 reports $\hat{\alpha}$ and $\hat{q}_\alpha$ defined in (17), with $\alpha = 0.01, 0.05$, together with the corresponding tests by He et al. (2023). For our hypothesis of interest, there is no evidence to reject the null for the two taxi series, but there is mild evidence (especially at the 1% level, with $\hat{\alpha}$ observed to be mildly larger than 1%) to conclude that, for the Fama–French time series, there is no Kronecker product structure along $\{1, 2\}$. In other words, there is evidence to suggest that the portfolio return series have structures deviating from the low-rank structure along their respective categorizations by market equity and book-to-equity ratio, implying that the vectorized data may have a more distinct factor structure.

Unsurprisingly, the evidence for rejecting the null hypothesis of a Khatri–Rao product structure when using the CP TFM for both portfolio time series is slightly stronger than that for rejecting the Kronecker product structure using the Tucker TFM. This is because, at both $\alpha = 1\%, 5\%$, the corresponding $\hat{q}_\alpha$ values under the CP TFM are larger than those under the Tucker TFM. This should be expected as the CP TFM using two factors here is a constrained version of the Tucker TFM using $(\hat{r}_1, \hat{r}_2) = (2, 2)$. Although our procedure does not directly test the choice between the Tucker TFM and CP TFM, it suggests using the Tucker TFM on a dataset if its Khatri–Rao product structure is rejected, but the Kronecker product structure is not. Otherwise, the CP TFM is preferred if neither structure is rejected, e.g., the results for the taxi series in Table 6.

Furthermore, the comparisons between the portfolio and residual series justifies the removal of the market effect, which is intuitive since the market effect should be pervasive in financial return data and is irrelevant to our categorizations.

For the image recognition series, the null is not rejected, which seems reasonable as the image alignment is important in recognizing digits. More interestingly, compared with all previous test results, our null is significantly rejected on the macroeconomic index data with $\hat{\alpha}$ clearly higher than the test levels. This suggests that the phenomenon observed in, e.g., Chen & Fan (2023) and He et al. (2023), where the estimated ranks of the matrix factor model fluctuate across different estimation methods, is likely due to pseudo ranks rather than insufficient cross-sectional dimensions. Lastly, for all datasets, we cannot reject the null by considering those alternative hypotheses considered in He et al. (2023).

6. Discussion

The factor models considered in Definition 2 are based on either a Tucker or CP decomposition, and this paper focuses on the former since CP decomposition is a special case of Tucker decomposition. Other tensor decompositions are possible, such as the low separation rank decomposition (Taki et al., 2024), which generalizes Tucker decomposition further. It is also easy to check that Theorem 1 holds for such a decomposition.

A more important remark is that, although this paper focuses on testing the Kronecker product structure in a Tucker TFM, our test is readily applicable to testing the Khatri–Rao product structure in a CP TFM; see the technical details in the Supplementary Material, particularly Appendix B therein. Although a direct extension of our test exists, it certainly merits developing more powerful tests by leveraging estimators specific to the CP TFM. This should also shed light on testing the CP TFM against the Tucker TFM, as demonstrated in the real data analyses, thereby addressing the problem of specifying the forms of tensor factor models.

Supplementary material

The Supplementary Material includes further details on the product notation, tensor reshaping and model identification, additional remarks, details of testing the Khatri–Rao product structure and proofs of all the theorems and auxiliary results. The testing procedures and tensor reshaping in this paper can be implemented with our `R` package `KOFM` available on CRAN (R Development Core Team, 2025).

REFERENCES

BARIGOZZI, M., HE, Y., LI, L. & TRAPANI, L. (2025). Statistical inference for large-dimensional tensor factor model by iterative projections. *arXiv*: 2206.09800v3.

CEN, Z. & LAM, C. (2025). Tensor time series imputation through tensor factor modelling. *J. Economet.* **249**, 105974. https://doi.org/10.1016/j.jeconom.2025.105974

CHANG, J., HE, J., YANG, L. & YAO, Q. (2023). Modelling matrix time series via a tensor CP-decomposition. *J. R. Statist. Soc. B* **85**, 127–48.

CHEN, E. Y. & FAN, J. (2023). Statistical inference for high-dimensional matrix-variate factor models. *J. Am. Statist. Assoc.* **118**, 1038–55.

CHEN, E. Y., XIA, D., CAI, C. & FAN, J. (2024). Semi-parametric tensor factor analysis by iteratively projected singular value decomposition. *J. R. Statist. Soc. B* **86**, 793–823.

CHEN, R., YANG, D. & ZHANG, C.-H. (2022). Factor models for high-dimensional tensor time series. *J. Am. Statist. Assoc.* **117**, 94–116.

CHEN, W. & LAM, C. (2024). Rank and factor loadings estimation in time series tensor factor model by pre-averaging. *Ann. Statist.* **52**, 364–391.

GUGGENBERGER, P., KLEIBERGEN, F. & MAVROEIDIS, S. (2023). A test for Kronecker product structure covariance matrix. *J. Economet.* **233**, 88–112.

HAN, Y., CHEN, R., YANG, D. & ZHANG, C.-H. (2024a). Tensor factor model estimation by iterative projection. *Ann. Statist.* **52**, 2641–67.

HAN, Y., YANG, D., ZHANG, C.-H. & CHEN, R. (2024b). CP factor model for dynamic tensors. *J. R. Statist. Soc. B* **86**, 1383–413.

HAN, Y., ZHANG, C. H. & CHEN, R. (2022). Rank determination in tensor factor model. *Electron. J. Statist.* **16**, 1726–803.

HE, Y., KONG, X., TRAPANI, L. & YU, L. (2023). One-way or two-way factor model for matrix sequences? *J. Economet.* **235**, 1981–2004.

HE, Y., WANG, Y., YU, L., ZHOU, W. & ZHOU, W.-X. (2022). Matrix Kendall's tau in high-dimensions: a robust statistic for matrix factor model. *arXiv*: 2207.09633v1.

LAM, C. & YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40**, 694–726.

ONATSKI, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *J. Economet.* **168**, 244–58.

R DEVELOPMENT CORE TEAM (2025). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.

TAKI, B., SARWATE, A. D. & BAJWA, W. U. (2024). Low separation rank in tensor generalized linear models: an asymptotic analysis. In *2024 58th Ann. Conf. Info. Sci. Syst. (CISS)*, pp. 1–6. Princeton, NJ: IEEE Press. https://dx.doi.org/10.1109/CISS59072.2024.10480200.

WANG, D., LIU, X. & CHEN, R. (2019). Factor models for matrix-valued high-dimensional time series. *J. Economet.* **208**, 231–48.

YU, L., HE, Y., KONG, X. & ZHANG, X. (2022a). Projected estimation for large-dimensional matrix factor models. *J. Economet.* **229**, 201–17.

YU, L., XIE, J. & ZHOU, W. (2022b). Testing kronecker product covariance matrices for high-dimensional matrix-variate data. *Biometrika* **110**, 799–814.

[*Received on* 21 *January* 2025. *Accepted on* 22 *September* 2025]